

PATENT
32860-000621/US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Frank KRICKHAHN

Application No.: NEW APPLICATION

Filed: September 30, 2003

For: **METHOD AND APPARATUS FOR STRUCTURING TEXTS**

PRIORITY LETTER

September 30, 2003

MAIL STOP NEW APPLICATION
COMMISSIONER FOR PATENTS
P.O. BOX 1450
Alexandria, Virginia 22313-1450

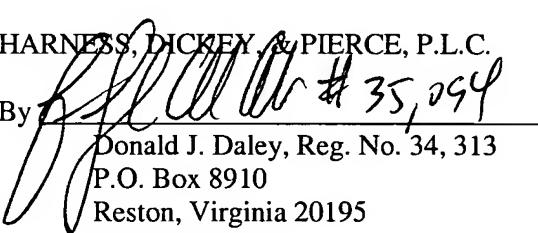
Dear Sirs:

Pursuant to the provisions of 35 U.S.C. 119, enclosed is/are a certified copy of the following priority document(s).

<u>Application No.</u>	<u>Date Filed</u>	<u>Country</u>
10245876.6	September 30, 2002	GERMANY

In support of Applicant's priority claim, please enter this document into the file.

Respectfully submitted,

HARNESS, DICKEY, & PIERCE, P.L.C.
By 
Donald J. Daley, Reg. No. 34,313
P.O. Box 8910
Reston, Virginia 20195
(703) 668-8000

DJD/JAC/jj

BUNDESREPUBLIK DEUTSCHLAND



Prioritätsbescheinigung über die Einreichung einer Patentanmeldung

Aktenzeichen: 102 45 876.6

Anmeldetag: 30. September 2002

Anmelder/Inhaber: Siemens Aktiengesellschaft, München/DE

Bezeichnung: Verfahren und Vorrichtung zur Strukturierung von Texten

IPC: G 06 F 17/27

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ursprünglichen Unterlagen dieser Patentanmeldung.

München, den 8. September 2003
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Stanschus

Beschreibung

Verfahren und Vorrichtung zur Strukturierung von Texten

5 Die Erfindung betrifft ein Verfahren und eine Vorrichtung zur Überführung unstrukturierter Textinformationen in ein strukturiertes Format.

10 Insbesondere in der Medizintechnik werden heutzutage viele Freitext-Reporte erstellt, die beispielsweise durch Einsatz von Diktaphonen und/oder Spracherkennungs-Technologien im Computer erfasst werden. Das Problem beim Umgang mit diesen Reporten besteht darin, dass ein automatischer Zugriff auf kleine Informationsteile, sogenannte atomare Informationen, 15 nahezu unmöglich ist, da der Inhalt keine bzw. nur eine sehr grobe Struktur enthält. Somit sind Freitext-Reporte für die strukturierte Darstellung und Auswertung der Informationen sehr ungeeignet.

20 Bei derartigen Freitext-Reporten werden lediglich ganzheitliche Informationen verarbeitet. Sie können nicht zu automatischen Auswertungen herangezogen werden, so dass die in ihnen enthaltenen Informationen somit für diesen Zweck verloren sind. Dieses Problem wächst in dem Maße, wie das Bedürfnis 25 nach Zugriff auf die atomare Information beispielsweise zum Zwecke eines Kodierens steigt.

Die Erfindung geht von der Aufgabe aus, ein Verfahren und eine Vorrichtung der eingangs genannten Art zu schaffen, die 30 eine einfache, automatisierte Überführung unstrukturierter Textinformationen von Freitext-Reporten in ein strukturiertes, auswertbares Format ermöglicht.

Die Aufgabe wird erfindungsgemäß durch ein Verfahren mit folgenden Schritten gelöst:

a) Eingabe von Strukturierungsregeln zur Strukturierung der unstrukturierten Textinformationen,

- b) Erfassen unstrukturierter Textinformationen,
- c) Parsen der unstrukturierten Textinformationen zur Erzeugung kleiner Textfragmenten,
- d) Durchsuchen von Texteinheiten der unstrukturierten Textinformationen nach in den Strukturierungsregeln definierten Textfragmenten,
- 5 e) Strukturierung der Textfragmente der unstrukturierten Textinformationen gemäß in den Strukturierungsregeln festgelegten Bedingungen.

10

Durch die zu definierenden Strukturierungsregeln wird der Freitext-Report geparsst, d.h. in kleinere Einheiten zerlegt, und in eine Struktur überführt, damit ein Programm diese Informationen auswerten kann. Eine solche Regel enthält Informationen zu den Textfragmenten, nach denen der Freitext-Report zu durchsuchen ist, welches Struktur-Element hierdurch repräsentiert wird, sowie zusätzliche Informationen darüber, wie die Struktur aufzubauen ist.

15

Erfindungsgemäß kann im Schritt b) die Erfassung von unstrukturierter Textinformationen durch ein Mikrofon erfolgen, wobei mittels eines Spracherkennungsprogramms eine Umsetzung in unstrukturierte Textinformationen durchgeführt wird.

25

In vorteilhafter Weise können die Strukturierungsregeln Informationen zu den Textfragmenten enthalten, nach denen der Freitext-Report zu durchsuchen ist, welches Struktur-Element hierdurch repräsentiert wird und wie die Struktur aufzubauen ist.

30

Die Aufgabe wird bezüglich der Vorrichtung erfindungsgemäß durch eine Eingabevorrichtung für unstrukturierte Textinformationen, eine Eingabevorrichtung und Speichervorrichtung für Strukturierungsregeln, eine Extraktionsvorrichtung von kleinen Texteinheiten aus den unstrukturierten Textinformationen, eine Strukturierungsvorrichtung zur Erzeugung strukturierter Textinformationen aufgrund der Strukturierungsregeln und eine

35

Auswertevorrichtung für die Texteinheiten in den strukturierten Textinformationen gelöst.

Eine Eingabe von auswertbaren unstrukturierten Textinformationen kann direkt erfolgen, wenn der Eingabevorrichtung für unstrukturierte Textinformationen eine Vorrichtung zur Spracherkennung zugeordnet ist.

Es hat sich als vorteilhaft erwiesen, wenn für die strukturierten Textinformationen als strukturiertes Format DICOM-SR oder XML Verwendung findet.

Die Erfindung ist nachfolgend anhand von in der Zeichnung dargestellten Ausführungsbeispielen näher erläutert. Es zeigen:

Figur 1 eine erfindungsgemäße Vorrichtung zur Strukturierung von Texten und

Figur 2 ein erfindungsgemäßes Verfahren zur Strukturierung von Texten.

In der Figur 1 ist eine erfindungsgemäße Vorrichtung zur Strukturierung von Texten dargestellt, die beispielsweise in einem Personal Computer (PC) realisiert sein kann. Eine Tastatur 1 dient zur Eingabe von Strukturierungsregeln und ggf. von Freitext-Reporten. Weiterhin kann die Vorrichtung eine Sprach-Eingabevorrichtung 2 aufweisen, beispielsweise ein Mikrophon oder ein Kassetten-Abspielgerät aufweisen, durch die die Freitext-Reporte in den PC eingegeben werden können. An die Sprach-Eingabevorrichtung 2 ist eine Vorrichtung 3 zur Spracherkennung, beispielsweise mit einem Spracherkennungsprogramm, angeschlossen, mittels derer die gesprochenen Freitext-Reporte in Textinformationen umgesetzt werden können.

Die Tastatur 1 ist mit einer Speichervorrichtung 4 für Strukturierungsregeln und einer Speichervorrichtung 5 für Textin-

formationen verbunden, an die auch die Vorrichtung 3 zur Spracherkennung angeschlossen ist. Mit der Speichervorrichtung 5 für Textinformationen ist eine Extraktionsvorrichtung 6 verbunden, die aus den unstrukturierten Textinformationen kleine Texteinheiten erkennt und kennzeichnet. An der Extraktionsvorrichtung 6 und der Speichervorrichtung 4 für die Strukturierungsregeln ist eine Strukturierungsvorrichtung 7 zur Erzeugung strukturierter Textinformationen angeschlossen, die aufgrund der festgelegten und gespeicherten Strukturierungsregeln die extrahierten Texteinheiten in ein strukturiertes Format überführt. Mit der Strukturierungsvorrichtung 7 ist eine Auswertevorrichtung 8 verbunden, die eine Abfrage nach kleinen, strukturierten Texteinheiten zur weiteren Auswertung ermöglicht.

15

In einer medizinischen Einrichtung werden Freitext-Reporte beispielsweise mit Hilfe eines Diktiergerätes (Dictaphones) aufgezeichnet und später von einer Sekretärin mittels eines Schreibprogramms über die Tastatur 1 in den Computer übertragen. Eine Umwandlung eines Freitext-Reports in einen geschriebenen Text kann auch durch die Vorrichtung 3 zur Spracherkennung mit einem entsprechenden Spracherkennungsprogramm erfolgen, wobei der Freitext-Report direkt in einen Personal Computer mittels Diktat oder nachträglich mit einem Abspielgerät für Diktierkassetten eingegeben werden kann.

25

Um spätere Auswertungen der so entstandenen Datenbestände zu ermöglichen, werden die Freitext-Reporte neben ihrem Originalformat in ein strukturiertes Format, beispielsweise DICOM-SR oder XML, überführt. Hierzu werden Regeln definiert, die die Systematik der Überführung festlegen.

30

Ausgangspunkt sind in der Figur 2 dargestellte unstrukturierte Textinformationen 9, welche mittels Diktat bzw. Freitexteingabe entstanden ist. Diese Textinformationen 9 dient als Input für eine Vorrichtung, die diese unstrukturierten Textinformationen 9 in eine strukturierte Form übersetzen soll.

In der Figur 2 ist Folgendes als Beispiel für unstrukturierte Textinformationen 9 angegeben:

5 Indikation: Diaphorese. Ausschluss von Abnormalitäten regionaler Wandbewegungen. Überprüfen hypertonischer Kardiomyopathie. Ausschluss myokardialen Infarkt. Beurteilen des linken des Auswurfanteils des linken Ventrikels. Ausschluss eines Aneurysma des linken Ventrikels.

10 Historie: Andere sachbezogene Historien beinhalten: neuerlicher Kokainmissbrauch. Vorhergehende CV-

Prozeduren:

Studieninfo. Die Studie wurde unter generaler Anästhesie durchgeführt.

15 Zur Übersetzung dieser unstrukturierten Textinformationen 9 in eine strukturierte Form werden in diese Vorrichtung Strukturierungsregeln 10 über die Tastatur 1 eingegeben und in der Speichervorrichtung 4 gespeichert, welche die Grundlage der Übersetzung bilden.

20 Diese Strukturierungsregeln 10 definieren, nach welchen Textfragmenten im Text zu suchen ist und welches Ergebnis das Auffinden eines solchen Textfragmentes in der Übersetzung hat. Im nachfolgend beschriebenen Beispiel bedeutet beispielweise das Auffinden des Textfragmentes "Indikation" oder "Indications", dass eine neues Element, welches eine Indikation beschreibt, in die Struktur eingefügt wird.

25 Im Folgenden sind Beispiele von derartigen Strukturierungsregeln 10 angegeben, die in der Figur 2 dargestellt sind. Die allgemeine Grundlage ist, dass Strukturierungsregeln 10 definiert werden, welche auf der Basis des Auffindens von Textfragmenten festlegen, wie unstrukturierte Textinformationen 9 in eine strukturierte Form übertragen werden.

30 Steht im Text das Wort "Indikation", so ist es unter Element "Indikation" mit offenen Aktionen zu behandeln. Gleiches gilt

für das Wort "Historie" als Element "Historie" und "Studieninfo" als Element "Studieninfo".

5 Steht im Text das Wort "Diaphorese", so ist es unter Element "Indikation" als Aktion einzufügen. Das Wort "Kokainmissbrauch" im Text ist unter Element "Historie-Eintrag" einzufügen. Der Begriff "generale Anästhesie" ist unter Element "Studieninfo" einzufügen

10 Mit diesen und weiteren einmal eingegebenen, jedoch jederzeit änderbaren Strukturierungsregeln 10 werden unstrukturierte Textinformationen 9 des Freitext-Reports in eine strukturier-
te Form gebracht, so dass man in den nun erhaltenen, nachfol-
gend beschriebenen strukturierten Textinformationen 11 nach
15 bestimmten Begriffen suchen kann.

```
<Report>
<Indikationen>
<Indikation> Diaphorese</ Indikation >. Ausschluss von
Abnormalitäten regionaler Wandbewegungen. Überprüfen hy-
20 pertonischer Kardiomyopathie. Ausschluss myokardialen
Infarkt. Beurteilen des linken des Auswurfanteils des
linken Ventrikels. Ausschluss eines Aneurysma des linken
Ventrikels.
```

```
</Indikationen>
```

```
<Historie>
Andere sachbezogene Historien beinhalten: neuerlicher
25 <Historie-Eintrag> Kokainmissbrauch <Historie-Eintrag>.
Vorhergehende CV-Prozedur(en):
```

```
</Historie>
```

```
30 <Studieninfos>
Die Studie wurde unter <Studieninfo> generaler Anästhe-
sie <Studieninfo> durchgeführt.
```

```
</Studieninfos>
```

```
</Report>
```

35 Es ist also ein automatischer Zugriff auf atomare Informatio-
nen möglich, da der Inhalt durch die erfundungsgemäße Vor-

richtung eine fein strukturierte Form erhält. Somit lassen sich auch Freitext-Reporte für eine strukturierte Darstellung und automatische Auswertung der Informationen einsetzen.

Patentansprüche

1. Verfahren zur regelbasierten Überführung unstrukturierter Textinformationen in ein strukturiertes Format mit folgenden

5 Schritten:

- a) Eingabe von Strukturierungsregeln (10) zur Strukturierung der unstrukturierten Textinformationen (9),
- b) Erfassen unstrukturierter Textinformationen (9),
- c) Parsen der unstrukturierten Textinformationen (9) zur Er-
10 zeugung kleiner Textfragmenten,
- d) Durchsuchen von Texteinheiten der unstrukturierten Textin-
formationen (9) nach in den Strukturierungsregeln (10) de-
finierten Textfragmenten,
- e) Strukturierung der Textfragmente der unstrukturierten
15 Textinformationen (9) gemäß in den Strukturierungsregeln
(10) festgelegten Bedingungen.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass im Schritt b) die Erfassung von unstrukturierter Textinformationen (9) durch ein Mikro-
20 phon erfolgt, wobei mittels eines Spracherkennungsprogramms eine Umsetzung in unstrukturierte Textinformationen durchge-
führt wird.

25 3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass die Strukturierungsre-
geln (10) Informationen zu den Textfragmenten enthalten, nach denen der Freitext-Report zu durchsuchen ist.

30 4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass die Strukturierungsregeln (10) Informationen zu den Textfragmen-
ten enthalten, welches Struktur-Element hierdurch repräsen-
tiert wird.

35 5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, dass die

Strukturierungsregeln (10) Informationen enthalten, wie die Struktur aufzubauen ist.

6. Vorrichtung zur regelbasierten Überführung unstrukturier-
5 ter Textinformationen in ein strukturiertes Format mit einer Eingabevorrichtung (1, 2) für unstrukturierte Textinformatio-
nen (9), mit einer Eingabevorrichtung (1) und Speichervor-
richtung (4) für Strukturierungsregeln (10), mit einer Ex-
traktionsvorrichtung (6) von kleinen Texteinheiten aus den
10 unstrukturierten Textinformationen, mit einer Strukturie-
rungsvorrichtung (7) zur Erzeugung strukturierter Textinfor-
mationen (11) aufgrund der Strukturierungsregeln (10) und mit
einer Auswertevorrichtung (8) für die Texteinheiten in den
strukturierten Textinformationen (11).

15 7. Vorrichtung nach Anspruch 6, durch ge-
kennzeichnet, dass der Eingabevorrichtung (2)
für unstrukturierte Textinformationen (9) eine Vorrichtung
(3) zur Spracherkennung zugeordnet ist.

20 8. Vorrichtung nach Anspruch 6 oder 7, durch
gekennzeichnet, dass für die strukturierten
Textinformationen (9) als strukturiertes Format DICOM-SR Ver-
wendung findet.

25 9. Vorrichtung nach einem der Ansprüche 6 bis 8, durch
gekennzeichnet, dass für die strukturierten Textinformationen (9) als strukturiertes For-
mat XML Verwendung findet.

FIG 1

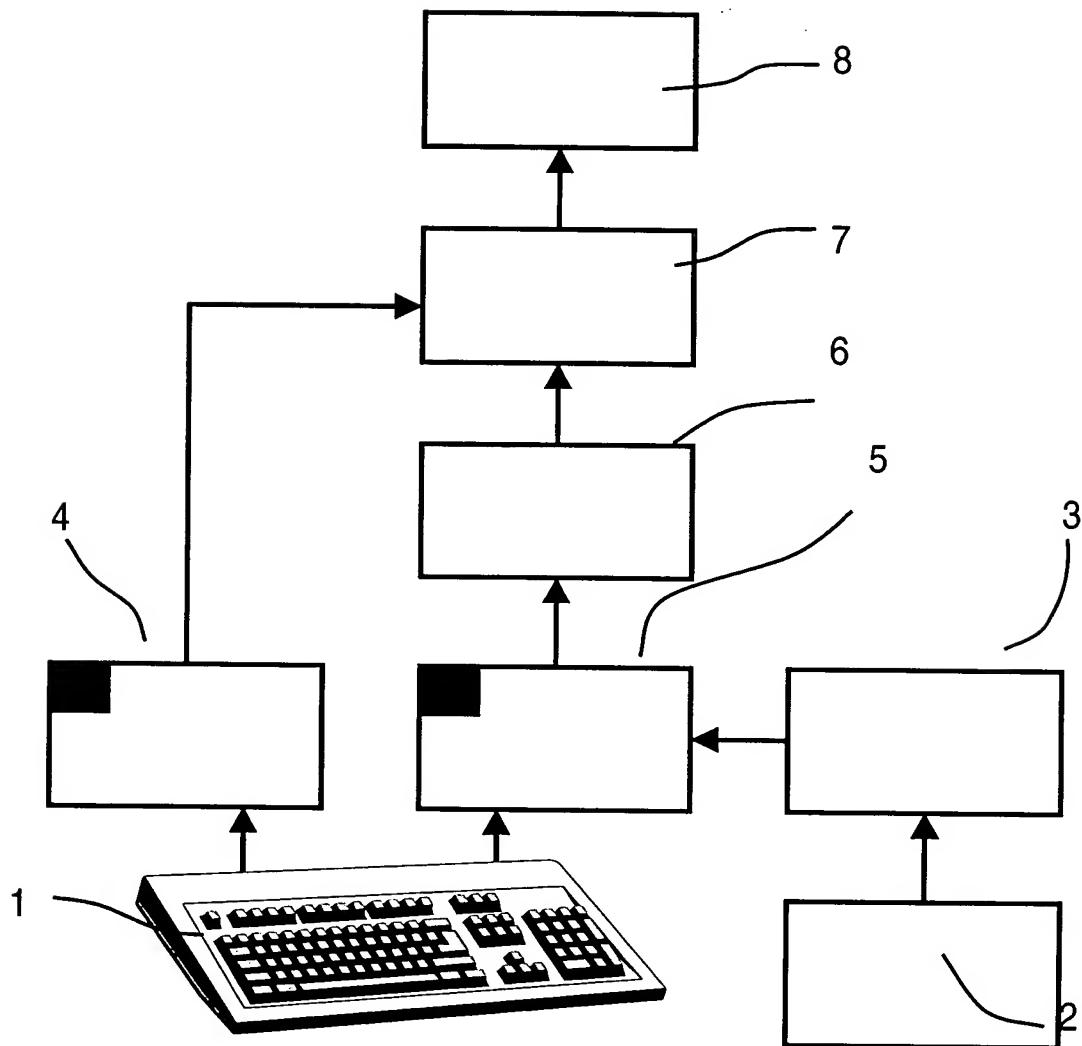


FIG 2

Indikation: Diaphorese. Ausschluss von Abnormalitäten regionaler Wandbewegungen. Überprüfen hypertonischer Kardiomyopathie. Ausschluss myokardialen Infarkt. Beurteilen des linken des Auswurfanteils des linken Ventrikels. Ausschluss eines Aneurysma des linken Ventrikels. Historie: Andere sachbezogene Historien beinhalten: neuerlicher Kokainmissbrauch. Vorhergehende CV-Prozeduren: Studieninfo. Die Studie wurde unter generaler Anästhesie durchgeführt.

9



Regel 1	Regel 2	Regel 3
Text = „Indikation“ Element = „Indikation“ Aktionen = „Offen“	Text = „Historie“ Element = „Historie“ Aktionen = „Offen“	Text = „Studieninfo“ Element = „Studieninfo“ Aktionen = „Offen“
Regel 4	Regel 1??	Regel 1??
Text = „ Diaphorese“ Element = „Indikation“ Aktionen = „Einfügung“	Text = „Kokainmissbrauch“ Element = „Historie-Eintrag“ Aktionen = „Einfügung“	Text = „generale Anästhesie“ Element = „Studieninfo“ Aktionen = „Einfügung“

10



```

<Report>
<Indikationen>
<Indikation> Diaphorese</ Indikation >. Ausschluss von Abnormalitäten regionaler Wand-
bewegungen. Überprüfen hypertonischer Kardiomyopathie. Ausschluss myokardialen
Infarkt. Beurteilen des linken des Auswurfanteils des linken Ventrikels. Ausschluss
eines Aneurysma des linken Ventrikels.
</Indikationen>
<Historie>
Andere sachbezogene Historien beinhalten: neuerlicher <Historie-Eintrag> Kokainmis-
brauch <Historie-Eintrag>. Vorhergehende CV-Prozedur(en):
</Historie>
<Studieninfos >
Die Studie wurde unter <Studieninfo> generaler Anästhesie <Studieninfo> durchgeführt.
</Studieninfos>
</Report>

```

11